PKU EECS

RSA 密码破译报告

冯古豪 2000013175

2022 年 6 月 16 日

目录

摘要

RSA 加密使我在通信中最常见的加密协议之一，它的加密原理主要是基于数论中的欧拉定理。尽管一般情况下，破解 RSA 密码是 NPC 问题，但是对于数据上有各种缺陷的 RSA 密码，我们依然能够破译。在这个项目中，我实现了多种 RSA 攻击，并最终破解了 17 组 RSA 加密的数据。

1 前言

在这个项目中，我主要实现的攻击方式主要包括：低加密指数攻击，低加密指数广播攻击，维纳攻击，共模攻击，Copper-Smith 攻击以及基于大整数分解的攻击。其中，我尝试的大整数分解方式主要包括：费马分解，Pollard-$p-1$ 分解，Pollard-ρ 分解。最终，我成功破译了 17 组 RSA 加密的数据。

2 RSA 加密原理

RSA 是一种非对称的加密协议，它是基于数论中的欧拉定理设计的。下面我们会给出 RSA 加密的具体流程和证明。RSA 加密算法首先生成出两个大素数 p, q，要加密的二进制数记为 m，公共模数 $n = pq$，公钥 e 为任意一个小于 $\phi(n)$ 的正整数。私钥 d 满足 $e \times d \equiv 1 \mod \phi(n)$，加密后我们发送的数字 $c = m^e \mod n$。在解密时，有 $m = c^d \mod n$，其中，公钥为 c 公开的，私钥 d 只有通信两方掌握。其中，c, d 分别又被成为加密指数和解密指数，m, c 分别被称为明文和密文。在一般的实际应用中，m, n 一般为 1024 位二进制数，公钥 d 一般取 65537。而在实验的数据中，m, n 全为 1024 位的二进制数，公钥 d 并未全都取值 65537。证明 RSA 加密算法只要证明解密算法的正确性，即：$m = c^d \mod n$ 即可。

Theorem 1 (Euler)**.** *a, n 为两个互素的正整数，则 $a^{\phi(n)} \equiv 1 \mod n$，其中 $\phi(n)$ 为欧拉函数。*

证明. $\{1, \cdots, n-1\}$ 在模 n 的乘法下构成群 G，考虑子群

$$H = \{x, (x, n) = 1\}$$

$\forall x, y \in H$，$(xy, n) = 1$，从而 H 是良定义，$|H| = \phi(n)$。因为 $a \in H$，所以根据拉格朗日定理，有：

$$|a||H| = \phi(n)$$

从而，$a^{\phi(n)}$ 为子群的单位元，所以有

$$a^{\phi(n)} \equiv 1 \mod n$$

□

Theorem 2 (RSA)**.** $m = c^d \mod n$

证明. 即证：$m = m^{ed} \mod n$。记 $cd = k\phi(n) + 1$，当 m 与 n 互素的时候，由欧拉定理，$m^{\phi(n)} = tn + 1$，从而 $m = m^{cd} \mod n$；当 m 与 n 不互素的时候，除掉公因数后，该式子依然成立。 □

3 针对低加密数的攻击

当 RSA 加密指数 e 很小的时候，RSA 的安全性就可能受到很大的威胁，针对 RSA 的低加密指数，存在一系列的攻击方式，在本次实验中，我实现了三种典型的攻击低加密指数的算法。

3.1 低加密指数攻击

低加密指数攻击主要是针对加密指数 e 很小的情形，一般为 2, 3。有 RSA 加密的流程，我可以得到 $m = \sqrt[e]{kn + c},\ k \in N$，因此我可以尝试通过枚举 k 的取值来破解 RSA。这种攻击算法的原理很简单，实际应用起来，算法的时间复杂度很大。由于在我的实验数据中 $m \sim \Theta(n)$，所以 $k \sim \Theta(n^{e-1})$，这样高时间复杂度的算法无法在有效的时间内解决任何一组加密数据。

3.2 低加密指数广播攻击

当我对于同一条消息采用不同的公共模数 n 和相同的加密指数 e 进行加密时，我可以使用另一种十分高效的攻击算法进行破译，即低加密指数广播攻击。

3.2.1 算法原理

低加密指数广播攻击主要是利用了数论中的中国剩余定理，在一组数据包含同一条消息，使用多个公共模数和同一加密指数加密后的一系列密文的情况下，破译密文。下面我将介绍并证明该攻击算法。

Theorem 3 (中国剩余定理)**.** *假设整数 $m_1, \cdots, m_n$ 两两互素，则对任意的整数 $a_1, \cdots, a_n$，对于以下方程组*

$$\begin{cases} x \equiv & a_1 \mod m_1 \\ x \equiv & a_2 \mod m_2 \\ & \cdots \\ x \equiv & a_n \mod m_n \end{cases}$$

记 $M = \Pi_{i=1}^{n} m_i$，$M_i = M/m_i$，t_i 为 M_i 模 m_i 下的倒数，则该方程组在模 M 的意义下有唯一解 $x = \Sigma_{i=1}^{n} a_i t_i M_i$

证明. 直接验证即可，对于任意的 i，$x \equiv a_i t_i M_i \mod m_i$，并且 $t_i M_i \equiv 1 \mod m_i$，所以

$$x \equiv a_i t_i M_i \equiv a_i \mod m_i$$

□

对于低加密指数广播攻击的情形，我们有：

$$\begin{cases} m^e \equiv & c_1 \mod n_1 \\ m^e \equiv & c_2 \mod n_2 \\ & \cdots \\ m^e \equiv & c_k \mod n_k \end{cases}$$

所以低加密指数广播攻击算法利用中国剩余定理可以得到：$m^e \equiv \Sigma_{i=1}^{k} c_i t_i N_i \mod N$，进一步，有 $m = \sqrt[e]{\Sigma_{i=1}^{k} c_i t_i N_i + KN}$。因此，该算法直接枚举整数 k 来完成破译。

3.2.2 算法分析和实验结果

由于实验数据基本上满足 $m \sim \Theta(n)$，所以枚举的 K 的范围接近 $\mathcal{O}(n^{e-k})$，为了得到有效的时间复杂度，该算法所需要数据的组数 k 需要满足 $k \geqslant e$。运用这个方法，我成功破译了 $e = 5$ 的五组数据。

3.3 Copper-Smith 攻击

这种攻击方式主要是针对解密指数 $e = 3$ 的情形，并且假设我们已知了 m 的高位数据。我们假设 $m = M + x$，其中，x 为我们要求解的低位数据。我们有

$$(M + x)^3 - c \equiv 0 \mod n$$

利用 Copper-Smith 算法，我们在较小的时间复杂度内解出满足条件 $x^3 - c \equiv 0 \mod n$ 的最小的 x，然后我们利用高位的数据就可以破译出 m。利用这种算法，并且结合数据的加密规则，我成功地破译了 $e = 3$ 的三组数据。

4 基于大数分解的攻击

对于 RSA 加密协议，如果我们能够分解公共模数 n 得到 p, q，这个时候，我们利用 $\phi(n) = (p - 1)(q - 1)$，可以计算出 $\phi(n)$，同时，知道 $\phi(n)$ 之后，我们就可以直接算出解密指数 d，如此就可以直接破译出 m。

4.1 费马分解

费马分解是一种常见的大数分解的算法，由于 $n = pq$，p, q 均为素数。不妨假设 $p \geqslant q$，这时，我们有 $n = (a - b)(a + b)$，$p = a + b$, $q = a + b$，从而有 $a^2 = n + b^2$。在这种思路下，费马分解有两种具体实现方式：

- $a = \sqrt[2]{n + b^2}$，我们可以枚举 b，来完成费马分解。
- $b = \sqrt[2]{a^2 - n}$，我们可以从 $a = \lceil \sqrt{n} \rceil$ 枚举，每次加一，来完成费马分解。

我们来比较两种方法枚举次数，假设后者枚举了 t 次，则 $b \sim O(\sqrt{(2t\sqrt{n} + t^2)})$。所以后者的枚举次数远远小于前者。在实验中，我分别尝试了两种实现方式，发现前者在有效的时间内只能够破译 1 组数据，但是后一种实现方式能够破译 3 组数据。

4.2 Pollard-$p - 1$ 解

Pollard-$p - 1$ 分解是一种很有效的大数分解算法，这个算法主要针对的是 $p - 1$ 包含的素因子都比较小的情况。

Lemma 4 (Pollard-$p - 1$)**.** *假设 $n = pq$ 满足 $(p - 1)|B!$ 且 p, q 均为素数，则令 $a = 2^{B!} \mod n$，$p = \gcd(a - 1, n)$。*

证明. 由于 $p|n$，所以 $a = 2^{B!} \mod p$，又由于 $(2, p) = 1$，根据费马小定理，$2^{p-1} \equiv 1 \mod p$，又由于 $(p - 1)|B!$，所以，$a \equiv 2^{B!} \equiv 2^{p-1} \equiv 1 \mod p$，从而 $p|(a - 1)$。由于 $a < n$，所以 $q \nmid a$，得证。 □

利用上述引理，Pollard-$p - 1$ 算法先设置一个较大的 B，然后计算出 $B!$, a，然后计算 $\gcd(a, n)$ 即可，若 $\gcd(a, n) \neq 1$，则我们解出了 n 的一个素因数 p，而后用 $q = n/p$ 即可分解 n。若 $\gcd(a, n) = 1$，则算法无效。在实验中，我是尝试了这种方法，我将 B 的大小设置成了 200000，而后针对每一组数据的 n 分别计算 a，最终成功破译了 3 组数据。

4.3 Pollard-ρ 解

Pollard-ρ 算法是一种比较有效的大数分解算法。这个算法的核心在于生成一个伪随机数列与 n 求解公因数，来分解 n。算法随机生成两个数 y, c，然后我们不断更新迭代 $y = f(y) = y^2 + c \mod n$，在 y 的迭代过程中，在模 n 的意义下是最终会构成一个环。

Lemma 5 (Pollard-ρ)**.** 如果 $|y_i - y_j| \equiv 0 \mod p$，则 $|y_{i+1} - y_{j+1}| \equiv 0 \mod p$。

证明.

$$\text{我们注意到 } |f(y_i) - f(y_j)| = |y_i + y_j| \times |y_i - y_j|,$$
$$\text{从而如果 } |y_i - y_j| \equiv 0 \mod p, \text{ 则 } |y_{i+1} - y_{j+1}| \equiv 0 \mod p。$$

□

根据上述引理，如果环上某一些距离 d 的两个点满足 $p \mid |y_i - y_j|$ 则环上任意两个距离为 d 的点满足要求，所以在实际应用时，我们只需要对每一个距离 d 做一次判定即可。该算法使用经典的 *Floyd* 算法来判定环，开始随机生成三个数 x, y, c，迭代过程中，x 每迭代一次，y 迭代两次，并计算 $|x - y| \mod n$ 与 n 公因数，直到找到 n 的非平凡因子。该算法的复杂度比较依赖于随机数生成器的随机性，算法期望复杂度为 $\mathcal{O}(n^{\frac{1}{4}} \log n)$。在实验中，Pollard-$\rho$ 算法成功破译了三组数据。

5 其他攻击

除了之前叙述的攻击方式之外，我还尝试一些其他的攻击方式，主要包括寻找公因数，共模攻击和维纳攻击

5.1 寻找质因数

这种攻击很简单，就是对于多组数据中的不同的 n，我们尝试将它们两两之间求最大公因数，若存在两组 n 之间的最大公因数不是 1，则这两组数据可以直接破译了。

5.2 共模攻击

主要针对两组数据中的明文 m 和公共模数 n 相同，但加密指数 e 互素的情况，我们分别记为 $e1$, $e2$。此时我们有

$$\begin{cases} m_1^e \equiv & c_1 \mod n \\ m_2^e \equiv & c_2 \mod n \end{cases}$$

Lemma 6 (共模攻击)**.** 记 s, t 满足 $se_1 + te_2 = 1$，则 $m \equiv c_1^s * c_2^t \mod n$

证明. $c_1^s c_2^t \equiv m^{se_1 + te_2} \equiv m \mod n$ □

从而基于上述引理，我们可以得到共模攻击算法，首先利用欧几里得算法计算出 s, t，计算出 $c_1^s c_2^t \mod n$ 即可。利用这个算法，我成功破译了两组数据。

方法		数据点
基于低加密指数的攻击	低加密指数攻击	无
	低加密指数广播攻击	4 5 7 13 18
	Copper-Smith 攻击	0 6 17
基于大数分解的攻击	费马分解	8 10
	Pollard-$p-1$ 分解	1 3 12
	Pollard-ρ 分解	3 12
其他攻击	公因数分解	2 19
	共模攻击	11 14
	维纳攻击	无

表 1: 实验结果

5.3 维纳攻击

维纳攻击主要是针对解密指数 d 比较小的情况，并且这种攻击算法还要求素因子 p, q 比较接近。这种攻击算法的核心在于使用连分数逼近的方法解出 $\phi(n)$，从而得到 p, q。这种攻击算法的理论基础，也就是为什么我们能够通过连分数逼近的方法得到 $\phi(n)$ 在于 *Legendre* 定理。

Definition 1 (连分数逼近)**.** 设 $a \in R$ 的连分数表示为 $[a_0, a_1, \cdots, a_n, \cdots]$，则 $\forall n$，$[a_0, a_1, \cdots, a_n]$ 为 a 的连分数逼近。

Theorem 7 (Legendre)**.** 若满足 $|\alpha - \frac{p}{q}| < \frac{1}{q^2}$，$(p, q) = 1$，则 $\frac{p}{q}$ 是 α 的连分数逼近。

基于勒让德定理，我们就可以得到维纳攻击。首先定义 $G = \gcd(p-1, q-1), \lambda(n) = \frac{\phi(n)}{G}$，我们有 $ed = \frac{K}{G}\phi(n) + 1$，其中，$K$ 为未知常数。将 K, G 约分得到 k, g，这样我们可以得到

$$\frac{e}{n} = \frac{k}{dg}\frac{\phi(n)}{n} + \frac{1}{dn}$$

这样我们有 $\frac{e}{n} \approx \frac{k}{dg}$，我们用 $\frac{e}{n}$ 的连分数展开逼近 $\frac{k}{dg}$，得到 dg, k 之后，带入之前的式子求得 $\phi(n)$，然后我们可以利用韦达定理解出 $n = pq$，这样就破译了 RSA 加密数据。

Theorem 8 (维纳攻击)**.** 当数据满足 $q < p < 2q$，$d < \frac{1}{3}n^{\frac{1}{4}}$ 的条件下，维纳攻击一定能够解出 p, q。

在本次实验的数据中，数据均没有满足维纳攻击的条件，所以我使用维纳攻击没有破译出任何数据。

6 实现细节和实验结果

在具体实现中，我主要使用 python 语言实现，主要利用了 gmpy2 的软件包进行大数运算，针对 Copper-Smith 攻击，我主要使用了 Sage-math 语言进行实现。在 Sage-math 中，这个软件包实现了许多数论的经典算法，我主要使用了其中的 Copper-Smith 算法。